Predicting Sector Index Movement with Microblogging Public Mood Time Series on Social Issues

Yujie Lu[1] **Jinlong Guo**[2] **Sakamoto Kotaro**[1] **Shibuki Hideyuki**[1] **Tatsunori Mori**[1]
[1]Graduate School of Environment and Information Science,
Yokohama National University, Yokohama, 2408501, JAPAN
[2]Graduate School of Library and Information Science,
University of Illinois at Urban-Champaign, Champaign, IL 61820, USA
{luyujie,sakamoto,shib,mori}@forest.eis.ynu.ac.jp,
jguo24@illinois.edu

Abstract

This paper develops a technique that unfolds public mood on social issues from real-time social media for sector index prediction. We first propose a low-dimensional support vector machine (SVM) classifier using surrounding information for twitter sentiment classification. Then, we generate public mood time series by aggregating message-level weighted daily mood (WDM) based on the sentiment classification results. Lastly, we evaluate our method against the real stock index in two kinds of time periods (fluctuating and monotonous) separately using static cross-correlation coefficient (CCF) and dynamic vector auto-regression (VAR). The experiments on "food safety" issue show that the proposed WDM method outperforms the word-level baseline method in predicting stock movement, especially during fluctuating period.

1 Introduction

Social media websites, such as Twitter and Facebook, have generated a great amount of public opinions on a variety of issues, especially hot events and emergencies. As a result, user-generated content has become a significant resource for exploring useful knowledge. In the use of public mood entailed in the real-time message streaming, researchers have proposed a wide range of applications, for example, election prediction (Andranik et al., 2009), anti-terrorism assistance (Cheong and Lee, 2009) and consumer confidence poll (O'Connor et al., 2010). In this paper, we use it for stock prediction.

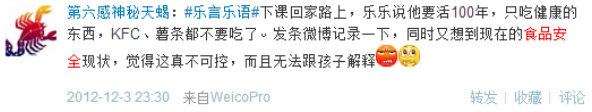

Figure 1: An example tweet from Sina Weibo

Sina Weibo is a Twitter-like microblogging service in China. Launched in 2009, it now has near 200 million monthly active users[1], which makes it the most dominant social networking service in China. Users discuss all kinds of social topics and express their opinions on the platform. As an example, food safety issue has become a prominent social problem and caused much concern in recent years in China. Figure 1 is an example tweet talking about food safety from Sina Weibo, in which the author expresses his dissatisfaction to the situation of food safety in China. Note that besides the text part, there is auxiliary information around the text (called surrounding information in this paper).

Previous work shows that indicators from real-time media can conceivably be used to predict changes for many economic indexes (Bollen et al. 2011), and behavioral finance theory suggests that public mood can drive stock market (Nofsinger, 2005).Hence, we construct public mood time series by analyzing millions of tweets in a time span to predict stock movement in the corresponding period.

Our main contributions are summarized as follows:

- We investigate how microblogging public mood on certain social issues relates to the stock movement of the relevant sector. In this study, we conduct an experiment on the topic of "food safety" using tweets from Sina Weibo and Shenzhen Stock Exchange (SZSE) Food & Beverage Index.

[1] According to 2015 first quarter financial results released by Weibo Corp.

29th Pacific Asia Conference on Language, Information and Computation pages 563 - 571
Shanghai, China, October 30 - November 1, 2015

- We utilize not only the text part of the tweet, but also the non-text part, namely surrounding information and user information, and show that both sentiment classification and public mood time series can be improved in use of it.
- We study how the methods perform for different types of periods of stock index. Both CCF and VAR evaluation show that public mood time series has better predictive power during fluctuating period than monotonous period.

To the best of our knowledge, this work is the first to predict sector stock index by public mood time series on social issues in Chinese microblogging.

2 Related Work

2.1 Stock Prediction with Social Media

With the popularity of real-time social media, stock market prediction based on microblog has attracted more and more attention. Past work can be roughly categorized into two classes depending on whether sentiment is used or not.

One class is sentiment-based methodology using general tweets. Bollen et al. (2011) generated seven different public mood time series using Opinion-Finder and Google-Profile of Mood States. Both Granger causality analysis with Dow Jones Industrial Average and a Self-Organizing Fuzzy Neural Network predictor showed that "Calm" dimension had the best predictive effect. Vu et al. (2012) experimented a Decision Tree classifier with different combinations of features to predict daily up and down movement of the stock price of tech companies. They proved that positive/negative sentiment, bullish/bearish orientation, and stock price change of three previous days are effective features. Si et al. (2013) proposed a topic-based method called continuous Dirichlet Process Mixture to learn subtopics, drew sentiment time series by aggregating opinion words over the topic chains. The VAR analysis with Standard & Poor's 100 showed its effectiveness.

The other class is non-sentiment-based methodology using financial tweets. Bar-Haim et al. (2011) distinguished expert users from non-experts according to the correctness of stock rise prediction against one's bullish posts. The precision of predicting stock rise showed that Per-User Model after expert classification performed better than other pattern methods. Ruiz et al. (2012) represented financial tweet sets as graphs, and extracted activity features and graph features. The correlation analysis with stock market activities showed that the number of connected components is the best feature, and the correlation with traded volume is stronger than stock price.

Our method belongs to the former. The main deference from previous work is that our public mood time series is based on message-level sentiment analysis on general tweets, and we creatively involve non-text information. Besides, unlike Bollen et al. (2011) predicting composite index value or Vu et al. (2012) forecasting individual company stock price, we observe how public mood on social issues affects stock movement at sector level.

2.2 Sentiment Analysis in Social Media

Pang et al. (2002) and Turney et al. (2002) are generally regarded as the start of the research area of sentiment analysis. These two works represent the two main methodologies of sentiment analysis — supervised method and unsupervised method. Pang fed machine learning methods, including support vector machine, maximum entropy, and Naïve Bayes, with features such as n-gram, part of speech to classify the polarity of texts. On the other hand, Turney calculated the comprehensive polarity of a text by aggregating the similarity between the keywords in the text and the seed words, which is known as SO-PMI algorithm. Broader overviews on traditional sentiment analysis are presented in Pang and Lee (2008) and Liu (2012).

Recent studies on sentiment analysis focus on social media. As an early attempt, Go et al. (2009) annotated a noisy training set based on emoticons in tweets, carried out analogous experiments as Pang et al. (2002), and showed that SVM classier achieved the best precision. Pak and Paroublek (2010) proposed a Naïve Bayes classifier using n-gram (embedded in POS distribution), and concluded that 2-gram worked the best. The SemEval Task reports (Nakov et al., 2013; Rosenthal et al., 2014) pointed out that participants leveraged various features, depended heavily on sentiment lexicons, and obtained the best accuracy around 70%. Xiang and Zhou (2014) proposed a topic-based sentiment mixture model, and achieved higher precision than the top systems in SemEval 2013.

Despite some special characteristics of Sina Weibo, sentiment analysis of Sina Weibo is similar to Twitter. Wang and Li (2014) proposed a SVM classifier with three-layered features which aggregate syno-

nyms and highly-related words to help reduce feature dimension, and indicated that it was better than SVM classifiers using n-gram and POS tags. Xie et al. (2012) proposed a set of weibo-specific features, such as the number of emoticon, for SVM classifier, and achieved an accuracy around 67%.

Concerning word-based features unavoidably cause data sparseness problem, similar to Xie et al. (2012), we use a SVM classifier with microblog-specific low-dimensional features due to its flexibility and efficiency. However, unlike previous work that only employs the text part of a tweet, we also make use of non-text information, such as the number of retweet and the number of reply.

3 Approach Outline

The overall framework of our research is shown in Figure 2.The core of our method is to build a sound public mood time series curve from tweets. This includes two main steps — bullish/bearish orientation representation and daily mood indicator design. Regarding the manifestation of bullish/bearish orientation, instead of using lexicon-based word sentiment of general tweets (Bollen et al., 2011; Si et al., 2013) or explicit buy/sell transaction of stock tweets (Bar-Haim et al., 2011), we utilize global polarity of general tweets, since global polarity contains more accurate emotion about its related object and general tweets allow us to have a wider base (Vu et al., 2012).In our study, tweets are divided into three categories: "positive", "negative" and "neutral". A positive tweet can be a potential "bullish" signal and a negative message can be a potential "bearish" signal for stock price.

To have a better message-level sentiment classification, we train a customized classifier for our selected topic instead of using existing general tools (e.g. OpinionFinder). We first extract text features and non-text features from tweets and feed the classier with different combination of them to find the best classifier. Using the customized classifier, we then obtain the global polarity of each tweet. Rather than using simple sentiment ratio as daily mood, we take non-text information into account to design a weighted daily mood indicator. The public mood time series curve can be easily drawn once we had weighted daily mood values of each day. We adopt two different perspectives to evaluate the prediction

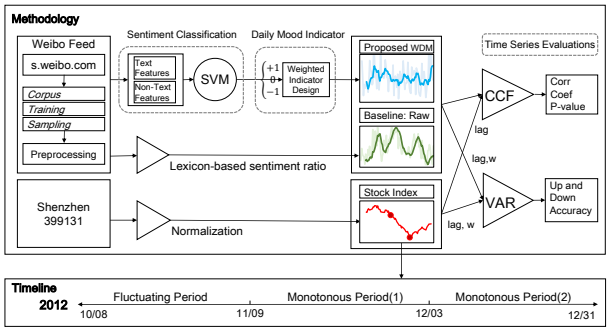

Figure 2: Overview of the research

ability of mood curves — CCF and VAR. Moreover, as shown at the bottom of Figure 2, the stock index is divided into fluctuating period and two monotonous periods according to the degree of volatility. We will compare how differently mood curves perform during the two kinds of time periods.

4 Customized Sentiment Classification

Both Pang et al. (2002) and Go et al. (2009) reported that SVM outperformed other classifiers where n-gram and POS features are used and unigram feature worked the best for traditional and twitter sentiment analysis respectively. Therefore, we choose SVM as our classier. Given the limited length of microblogging (only 140 characters), word-based n-gram and POS features lead to severe sparseness problem, so we design our microblog-specific features for SVM classifier.

4.1 SVM Classifier

Support Vector Machine (SVM) has proved to be an efficient classification model. The basic idea of it is to find a hyperplane represented by its normal vector $\mathbf{w}$ which maximizes the margin (the distance from the closest instances).This search then becomes a constrained optimization problem and the solution can be written as:

$$\mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i x_i, \alpha_i \geq 0 \qquad (1)$$

where α_i can be obtained by standard quadratic programming, x_i are support vectors lying on the margin and $y_i \in \{1, -1\}$[2]. SVM can solve non-linear tasks using kernel trick as well. In this work, our SVM classifier is trained with LibSVM toolkit using RBF kernel (Chang and Lin, 2011).

[2] Multiclass classification problem can be turned into multiple binary classification.

4.2 Text Features

Besides traditional text features like n-gram, POS tags and simple lexicon-based emotional word number, there are many microblogging-specific features in the text part of the tweet.

Entity Tag Count Entities are special elements in microblog. We exploit four kinds of often-used entities: hashtag, @tag, URL and seed. The former three are the same as Twitter, while the last one is a weibo-specific entity which allows users to subscribe RSS news about tagged words. The number of the four kinds of entities are used as features. These features were also used in previous work.

Set-count Neutral Signals Based on observation of many tweets, we collect heuristic neutral signals for identifying objective tweets. The more neutral signals a tweet contains, the more possible it is objective. The neutral signals consist of two subsets. One subset includes: bracket pair (【】), book title mark (《》), time patterns(e.g. *月*日) and numbers(e.g. 35%), and the other contains 5 types of words: news vocabulary (e.g. 宣传日), Q&A words (e.g. 科普), stock terms (e.g. 沪指), sharing words (e.g. 下载), and irrelevant words (e.g. 抽奖). Neutral signals are set-count features[3], so there are two features in total.

Sentence Count Unlike English tweet, Chinese tweet can easily have 3 or more sentences, so sentence information is important for weibo. We count the number of sentences, the number of exclamatory sentence indicated by exclamation marks, and the number of questions indicated by question marks.

The sum of emotional words is the basic element to measure the sentiment of a sentence or a message. We compute sentiment scores at both sentence and message level. They are defined as:

$$\text{Score(U)} = \sum_{i=1}^{|U|} \text{polarity(i)} \qquad (2)$$

where U denotes a unit of text and i denotes a word or an emoticon whose polarity is in {1, 0,-1}.

Sentence Sentiment Score The first sentence and the last sentence are always more important than others. Thus we compute sentiment scores of them respectively. Firstly, we clear up tags (entities, emoticon etc.) in the raw tweet, normalize the abnormal full stops, tokenize the cleaned text using NLPIR/ICTCLAS and segment it into sentences by punctuation (period, semicolon, exclamation mark, question mark, and suspension points).Then we turn sentences into word polarity vectors and gain the sentence score by summing up all the values in the vector. For example, “各种|食品安全|问题|集中|爆发|，|有些|是|问题|，|有些|是|误解|。” is transformed to [0,0,-1,1,-1,0,0,0,-1,0,0,0,-1,0].This calculation relies heavily on the quality of polarity dictionaries. There are three open-source sentiment lexicons for Chinese: How-net dictionary, DTU ontology dictionary and NTU dictionary. By comparing the effectiveness of these lexicons and their combinations on a small test set, we use the integration of all of them.

Message Sentiment Score We compute two global sentiment scores by emoticons and emotional words respectively. Emoticon is such a special reference for noisy labeling (Pak and Paroubek, 2010) and a strong indicator of global polarity (Kouloumpis et al., 2011) that we consider it separately. Unlike the emoticons in English that are combinations of ASCII characters, Sina Weibo emoticons are stipulated icons. Thus, we first classified 72 often-used emoticons in Sina Weibo into 3 categories (positive, negative and neutral), then sum up their polarities as the global sentiment score. There are two emoticons at the end of the example tweet (see Figure 1). Global sentiment score by emotional words is computed the same as the sentence sentiment score but on a larger scale.

4.3 Non-Text Features

Apart from text features, there are many metadata of the tweet (surrounding information) and the author (user information). Previous studies have not made full use of these data. Since raw data is stored in HTML pages, basic fields enclosed by HTML tags can be extracted by HTML parser. We extract message ID, user ID, user badge, user nickname, sending date, sending source, the number of retweet, the number of replies, the state of embedded picture and video. Some of these fields are just identificati-

[3] A set-count feature is a count of the number of instances from a set of terms.

on with little meaning such as message ID and user ID, while other fields can potentially be useful features.

Surrounding Information Surrounding information refers to the fields surrounding the text part of the tweet (see Figure 1). In our study, user badge, the number of retweet, the number of replies, the state of embedded picture and video are selected as features.

User Information We can access the user information using Sina Weibo user interface by user ID. Many fields such as gender, city, badge, and brief introduction about the user can be returned. We only make use of three numeric fields: the number of follower, following and posted tweets.

5 Daily Mood Indicator Design

Bollen et al. (2011) has shown that daily #positive/ #negative ratio (happiness) time series can represent public mood and emotionally responded to hot social events. Different from Bollen's curves based on word polarity aggregation, our time series are built on message-level sentiment analysis.

Considering the sentiment distribution of our experiment topic is skewed at the message level (very few positive tweets on food safety problem), we use Eq.3 as our basic daily mood indicator instead. It also means the degree of happiness and is monotonically decreasing (the more there are negative tweets, the less it will be).The public mood of day t (denoted as Daily Mood, DM) is defined as:

$$\mathrm{DM(t)} = \frac{\#_{\mathrm{t}}(\mathrm{tweet})}{\#_{\mathrm{t}}(\mathrm{tweet\_})} \tag{3}$$

where $\#_t(\mathrm{tweet})$ denotes the number of tweets in date t and $\#_t(\mathrm{tweet\_})$ denotes the number of negative tweets in date t.

Different tweets have different weights. A tweet that has many retweets or posted by famous people will have stronger impact on public mood and then on stock market. So we need to take these useful non-text fields into account. The weighted daily mood (WDM) and $\mathrm{Weight(t)}$ are represented as:

$$\mathrm{WDM(t)} = \mathrm{DM(t)} * \mathrm{Weight(t)} \tag{4}$$

$$\mathrm{Weight(t)} = \log_2\left(\frac{\sum_{\mathrm{t}}(\mathrm{retweet})}{\sum_{\mathrm{t,-}}(\mathrm{retweet})} * \frac{\lg(\sum_{\mathrm{t}}(\mathrm{follower}))}{\lg(\sum_{\mathrm{t,-}}(\mathrm{follower}))}\right) \tag{5}$$

where retweet means the number of the retweets of a tweet and follower means the number of the followers of the author of the tweet. We compute the total number of them in day t. Since follower is much greater than retweet, it is log transformed. The product is also log transformed for order reduction.

6 Experiment on Sentiment Classification

6.1 Text Data

Given that Sina Weibo API does not provide search interface freely as Twitter does, we scrape tweets discussing food safety with the keyword — "食品安全" (food safety in Chinese) from its search service platform[4]. Unlike Twitter API, Sina Weibo search platform allows to backtrack until 2009.The collecting interval is the fourth season of 2012(Oct. 1st 2012- Dec.31st 2012) when food safety problem was the most concerned problem for Chinese people. To sidestep undesired repetition, the original option is ticked. Totally we fetched 51,611 pieces of tweets (denoted as *Corpus*).

A training dataset is annotated for SVM classifier (denoted as *Training*). In accordance with Go et al. (2009), the definition of our polarity is "a personal positive or negative feeling". The polarity is presented as "+1(positive), 0(neutral), and -1(negative)". In addition, irrelevant tweets and objective tweets (e.g. news, commercial) are regarded as 0(as strict neutral ones consisting of both positive and negative are rare). All the tweets were tagged with one of {+1, 0,-1} by annotators. *Training* consists of 901 pieces of labelled messages coming from a randomly selected date.

6.2 Sector Index

In order to evaluate our public mood time series, a sector stock index for food industry is needed. We select SZSE Food & Beverage Index 399131(denoted as *Index*) as our stock index. *Index* consists of 56 main companies in food sector of China. The pe-

[4] http://s.weibo.com/

riod of *Index* corresponds to *Corpus* collecting period (Oct. 1st 2012- Dec.31st 2012)[5].To make it continuous, the values at weekends is computed by linear interpolation[6].

Figure 3 shows the *Index* curve (in order to compare with mood curves, the curve is Z-score normalized). As we can see, there are continuous decline and increase periods in the curve. On one hand, these long-term (soft) monotonous movement will render prediction more difficult since public mood changes drastically. On the other hand, prediction in long-term monotonous periods in less meaningful than it is in fluctuating periods for stock investors. So we discuss prediction in two types of periods: fluctuating period (Oct.8th - Nov.9th) and monotonous periods (Nov.10th - Dec.3rd, and Dec 4th - Dec.31st).

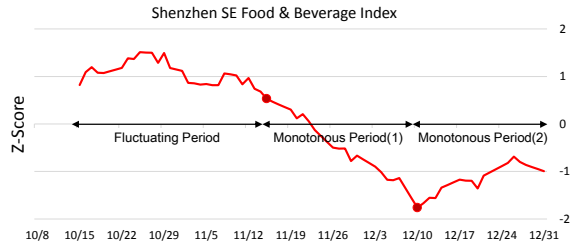

Figure 3: SZSE Food & Beverage closing values (Oct. 8th- Dec.31st)

6.3 Classification Result

The unigram method for sentiment classification described in Go et al. (2009) is used as a baseline. We employ WEKA[7] to construct the unigram model and classify tweets by its embedded LibSVM. We tried three combinations of our features. The evaluation method is five-fold cross-validation. Table 2 show the precision of each method.

Features	#Dim	Precision
baseline: unigram	2517	79.69%
C1:text features only	13	89.79%
C2:C1 + surrounding info	17	92.23%
C3:C2 + user info	20	87.35%

Table 1: The results of different classifiers

From Table 2 we can see:

1. C* classifiers perform better than the baseline by 10.1% on average. In addition, the number of the dimension of C* classifiers is far way less than baseline, which saves learning time. The result also implies that the traditional classification methods based on words have limitation for sentiment analysis, because word alone is not necessarily the carrier of emotion. Hence, although the dimension is very high, each of them does not contribute much. In contrast, each of our features has its underlying influence on the global polarity.

2. C2 is higher than C1 by 2.44%, and C3 decreased by 4.88%.This suggests surrounding information improves the classification, while user information does not. This makes sense because we know controversial tweets on social issues having many retweets or replies are more likely to be emotional. On the contrary, user information is not only different from other features in magnitude, but also incompatible with them in quality so that it disturbs the learning. This indicates that message sentiment is mainly decided by tweet text and its surrounding information.

As a result, we utilize C2 as our model. Now we look into the precision of different categories. The precision for neutral class reaches an impressive 98%, for negative class (majority) reaches 72.3%, both of which are higher than Xie et al. (2012) [8]. However, public mood on social events always goes to extremes. The majority of subjective class in *Corpus* is negative, because public mood for food safety in China is irritated at the collecting period. There are only 8 positive samples in *Training* and only 1 of them are classified correctly. Consequently, the prediction for positive tweets is unreliable. In fact, according to manual check, the positive tweets account for less than 1% of *Corpus*. This is why we changed the definition of daily mood in Section 5.

6.4 SVM Mood Curve & Sample Mood Curve

In theory, we simulate the real mood curve based on the result of SVM classier, but what if the real mood curve itself has no predictive power at the first place? In order to make sure whether there is a relationship between the real mood curve and the stock index curve, we annotated another larger dataset (denoted as *Sample*). *Sample* is sampled from tweets in *Corpus* during fluctuating period (Oct.8th-Nov.9th) at the

[5] Unfortunately 399131 *Index* has been delisted from Mar 1st, 2013.

[6] Since Oct 1st – Oct 7th is national holiday in China, we ignore these days.

[7] http://www.cs.waikato.ac.nz/ml/weka/

[8] This is a loose comparison because the training dataset is different.

rate of 20% (4106 tweets in total)[9]. Each tweet has been tagged by two independent annotators, and the agreement rate between annotators is 88%. The organizer double-check the left inconsistent 12%, and decide the final polarity.

First, we see how close SVM-based curve is to *Sample*-based curve. Figure 4 shows the two curves. The vertical axis is WDM value. Figure 4 suggests that the two curves are correlated significantly (p-value of correlation analysis < 0.01), which means that the C2 classifier is reliable for building WDM time series. The prediction performance of *Sample*-based curve is shown in the next section.

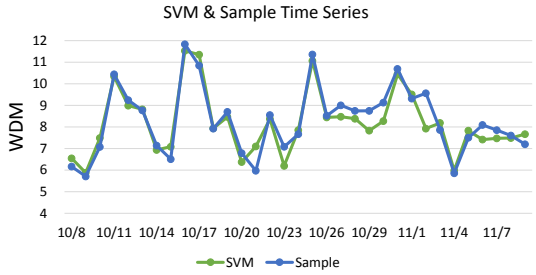

Figure 4: SVM & Sample WDM Time Series (Oct.8th - Nov.9th)

7 Experiment on Mood Time Series

Stock prediction is an extremely complex process. To better verify the prediction effect of proposed mood time series, we evaluate it in two ways (CCF and VAR). CCF observes the static similarity between mood time series and stock index, while VAR assesses the dynamic one-day-ahead prediction ability of mood time series. Besides, we evaluate the proposed method separately during fluctuating period and monotonous periods.

7.1 Public Mood Time Series

We apply the best C2 model to predict the polarity for each tweet in *Corpus*. Since there is not yet similar work on message-level sentiment time series, we use Bollen's method as our baseline (denoted as **Raw**).

Based on WDM, we can draw our proposed time series (denoted as **WDM**). For comparison, we also draw the DM time series ((denoted as **DM**)) and *Sample*-based mood time series (denoted as **Sample**)[10]. However, concerning that original public mood is highly vibrant (O'Connor et al. 2010), we smooth the mood curves by moving average over a window of the past 7 days. Smoothed time series of **Raw**, **DM**, and **WDM** are shown in Figure 5(Z-score normalized).

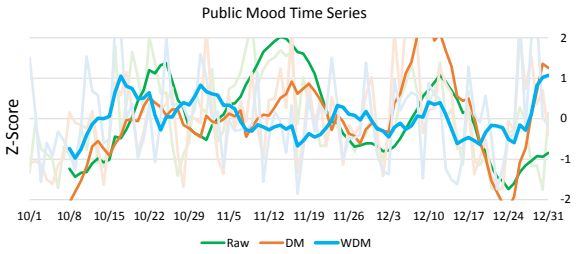

Figure 5: Public mood time series (Oct.8t -Dec.31st)

7.2 Cross-Correlation Coefficient

Cross-correlation coefficient shifts one curve back and forth to estimate correlation between two series at different time lag (Eduardo et al. 2012). We shift *Index* curve, so the right part where lag is greater than 0 means the ability to predict.

Figure 6 shows correlation coefficients between mood curves[t] and *Index* [t + lag]. We can see that the **WDM** curve has the best similarity with *Index* in prediction part in all the time spans. The average correlation value for **WDM** is 0.31 at predicting stage in entire period[11]. As expected, **WDM** has a similar trend with **Sample**, and what surprised us is that **WDM** is even higher than **Sample** curve. This may be because that **Sample** only contains 20% of *Corpus*, while **WDM** observes the whole *Corpus*. Moreover, It is obvious that **WDM** works better than simple **DM**， which verifies our idea that non-text information helps. Besides, we can see that **WDM** works much better in fluctuating period than monotonous periods and achieves the best value when lag is 2 in fluctuating period. On the other hand, both **DM** and **Raw** have little predictive ability in fluctuating period.

7.3 Vector Auto Regression

To access dynamic prediction ability, we use the vector auto-regression evaluation proposed in Si et al. (2013). The first order (lag=1) VAR model is defined as:

[9] The best way to obtain the real curve is to tag all the tweets in *Corpus*, but that is too large for manual annotation.

[10] To compare with **Sample**, the first 6 days of fluctuating period are cut off because of smoothing.

[11] Eduardo et al. (2012) reported 0.1 averagely on their time series.

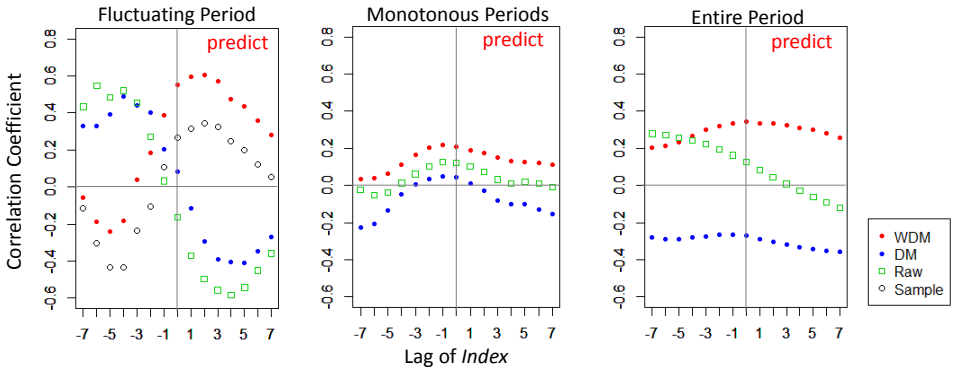

Figure 6: Correlation coefficient for different lags in different periods

$$x_t = \vartheta_{11} x_{t-1} + \vartheta_{12} y_{t-1} + \varepsilon_{x,t}$$
$$y_t = \vartheta_{21} x_{t-1} + \vartheta_{22} y_{t-1} + \varepsilon_{y,t} \quad (6)$$

The training data is a sliding window of the past w days. VAR uses the training data to predict the one-day-ahead up and down of *Index*. In our study, lag is in {1, 2, 3} and w is in {5, 10, 15}.Apart from mood curves, we test *Index* itself by univariate auto-regression model for reference. All curves are normalized to [0, 1].

Table 3 shows the average accuracy of the prediction in different lags. We can see from Table 3 that **WDM** performs best on average in fluctuating period, and achieves the highest accuracy 72.9% on lag 2, which is in accordance with the CCF result. Since the curve fluctuates much in this period, accuracy of *Index* itself is only 51.4%, which is nearly guess. However, if we look at the monotonous periods, all the three mood curves are worse than the *Index* itself. This is because the tendency in monotonous periods is very clear, *Index* itself can be a very strong predictor. Besides, **DM** performs the best among the mood curves. In the entire period, we combine a **W&D** curve using **WDM** in fluctuating period and **DM** in monotonous periods and achieves an accuracy of 65.3% averagely, performs better than **DM** or **WDM** alone. Since the monotonous periods is nearly twice the length as the fluctuating period, the overall accuracy does not win *Index*.

Fluctuating Period					
Lag	*Index*	Raw	DM	WDM	Hand
1	0.579	0.592	0.592	0.601	0.592
2	0.454	0.617	0.626	0.729	0.647
3	0.510	0.626	0.550	0.610	0.626
avg	0.514	0.612	0.589	**0.647**	0.622
Monotonous Periods					
Lag	*Index*	Raw	DM	WDM	
1	0.755	0.735	0.769	0.683	
2	0.757	0.688	0.717	0.738	
3	0.797	0.695	0.683	0.667	
avg	0.769	0.706	**0.723**	0.696	
Entire Period					
Lag	*Index*	Raw	DM	WDM	W&D
1	0.694	0.653	0.658	0.636	0.673
2	0.677	0.668	0.659	0.683	0.678
3	0.685	0.600	0.634	0.591	0.606
avg	0.685	0.640	0.650	0.637	**0.653**

Table 2: Average accuracies over all training windows size and different lags in different periods (Boldface: best performance)

8 Conclusion and Future Work

In this paper, we presented a framework using public mood on social issues to predict sector index movement. We developed a low-dimensional supervised sentiment classifier and designed a weighted daily mood indicator.

We found non-text information of tweet was useful for both sentiment classification and daily mood design. Experiment results showed that our proposed method worked best in terms of static CCF. For predicting one-day-ahead up and down by VAR, mood curves perform better during fluctuating period.

Although we presented an experiment on the topic of "food safety", the described technique can be extended to any other social topics. In the future, we plan to experiment controversial topics, such as "genetically modified food". In addition, since the prediction power depends on period type, it's meaningful to judge where the boundary of the period types is. These will be part of our future work.

References

Alec Go, Richa Bhayani, and Lei Huang.2009. Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, pages 1–6.

Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), pages 1320–1326.

Bing Liu. 2012. Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1):1–167.

Bing Xiang and Liang Zhou.2014. Improving twitter sentiment analysis with topic-based mixture modeling and semi-supervised training. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL'14), pages 434–439.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. Foundations and trends in information retrieval, 2(1-2):1–135.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'02), pages 79–86.

Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. 2010. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM'10), pages 122–129.

Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. In ACM Transactions on Intelligent Systems and Technology, 2(3):27.

Eduardo J. Ruiz, Vagelis Hristidis, Carlos Castillo, Aristides Gionis, Alejandro Jaimes. 2012. Correlating financial time series with micro-blogging activity. In Proceedings of the fifth ACM international conference on Web search and data mining (WSDM'12), pages 513–522.

Efthymios Kouloumpis, Theresa Wilson, and Johanna Moore .2011.Twitter sentiment analysis: the good the bad and the OMG!. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM'11), pages 538–541.

Jianfeng Si, Arjun Mukherjee, Bing Liu, Qing Li, Huayi Li, and Xiaotie Deng. 2013. Exploiting topic based twitter sentiment for stock prediction. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL'13), pages 24–29.

Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011. Twitter mood predicts the stock market. Journal of Computational Science, 2(1):1–8.

John R. Nofsinger. 2005. Social Mood and Financial Economics. Journal of Behavioral Finance, 6(3):144–160.

Marc Cheong and Vincent C. S. Lee. 2011. A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter. Information Systems Frontiers, 13 (1):45–59.

Peter D. Turney.2002.Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In Proceedings of the Association for Computational Linguistics (ACL'02), pages 417-424.

Preslav Nakov, Sara Rosenthal, Zornitsa Kozareva, Veselin Stoyanov, Alan Ritter, and Theresa Wilson .2013.SemEval-2013 Task 2: Sentiment Analysis in Twitter. In Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval'13), pages 312–320.

Roy Bar-Haim, Elad Dinur, Ronen Feldman, Moshe Fresko, and Guy Goldstein.2011.Identifying and Following Expert Investors in Stock Microblogs. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP'11), pages 1310–1319.

Sara Rosenthal, Alan Ritter, Preslav Nakov, and Veselin Stoyanov. 2014. SemEval-2014 Task 9: Sentiment Analysis in Twitter. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval'14), pages 73–80.

Tien Thanh Vu, Shu Chang, Quang Thuy Ha, and Ni gel Collier. 2012. An Experiment in Integrating Sentiment Features for Tech Stock Prediction in Twitter. In Proceedings of the Workshop on Information Extraction and Entity Analytics on Social Media Data (COLING'12), pages 23–38.

Tumasjan Andranik, Sprenger Timm O, Sandner Philipp G, Welpe, and Isabell M. 2011. Election Forecasts with Twitter: How 140 Characters Reflect the Political Landscape. Social Science Computer Review, 29(4):402–418.

Xie Lixing, Zhou Ming, and Sun Maosong. 2012. Hierarchical structure based hybrid approach to sentiment analysis of Chinese microblog and its feature extraction. Journal of Chinese Information Processing, 26(1):73–83.